

AFRL-RI-RS-TM-2008-28
Final Technical Memorandum
October 2008



NEEDLE IN THE HAYSTACK SECURE COMMUNICATION

University of Central Florida

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RI-RS-TM-2008-28 HAS BEEN REVIEWED AND IS APPROVED FOR
PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/

DAVID H. HUGHES
Work Unit Manager

/s/

WARREN H. DEBANY, Jr.
Technical Advisor, Information Grid Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) OCT 08			2. REPORT TYPE Final		3. DATES COVERED (From - To) Jul 07 – Jul 08	
4. TITLE AND SUBTITLE NEEDLE IN THE HAYSTACK SECURE COMMUNICATION				5a. CONTRACT NUMBER N/A		
				5b. GRANT NUMBER FA8750-07-1-0203		
				5c. PROGRAM ELEMENT NUMBER 62702F		
6. AUTHOR(S) Guifang Li				5d. PROJECT NUMBER 66SG		
				5e. TASK NUMBER UC		
				5f. WORK UNIT NUMBER F1		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Central Florida Office of Research & Commercialization 4000 Central Florida Blvd. Orlando, FL 32816-8005					8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFRL/RIGE 525 Brooks Rd. Rome NY 13441-4505					10. SPONSOR/MONITOR'S ACRONYM(S) N/A	
					11. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-RI-RS-TM-2008-28	
12. DISTRIBUTION AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. 88ABW-2008-PA No.-0324						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT Needle in-Haystack (NIH) is a novel physical layer secure optical data encryption scheme is investigated. NIH hides information in classical generated noise in such a way that it is indistinguishable from that noise, unless a key is possessed by a legitimate user allowing extraction of the information.						
15. SUBJECT TERMS Optical data encryption, optical communication, cryptography.						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 12	19a. NAME OF RESPONSIBLE PERSON David H. Hughes	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) N/A	

1. Introduction

In conventional digital encryption, a plaintext message is first encoded using a key, for example, through permutation and substitution, to obtain a ciphertext. The ciphertext can only be decoded if the recipient has the same key used for encoding so that the reverse permutation and substitution can be made to recover the plaintext as shown in Fig. 1.

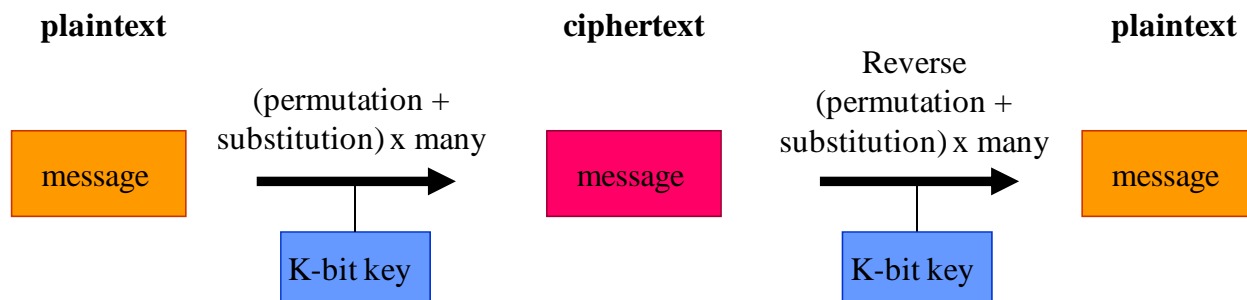


Figure 1 Schematic of conventional digital encryption.

Good encryption codes cannot be reliant on design being kept secret and must be resistant to known plaintext attacks. For a K -bit key, the algorithm requires brute force to break it in 2^K attempts. Encryption algorithm must be kept simple with reasonably short key to be cost effectively implemented using digital hardware. However, the consequence of simplicity is that a message might be breakable in the near future, e.g. 10 years, with advanced in computing power.

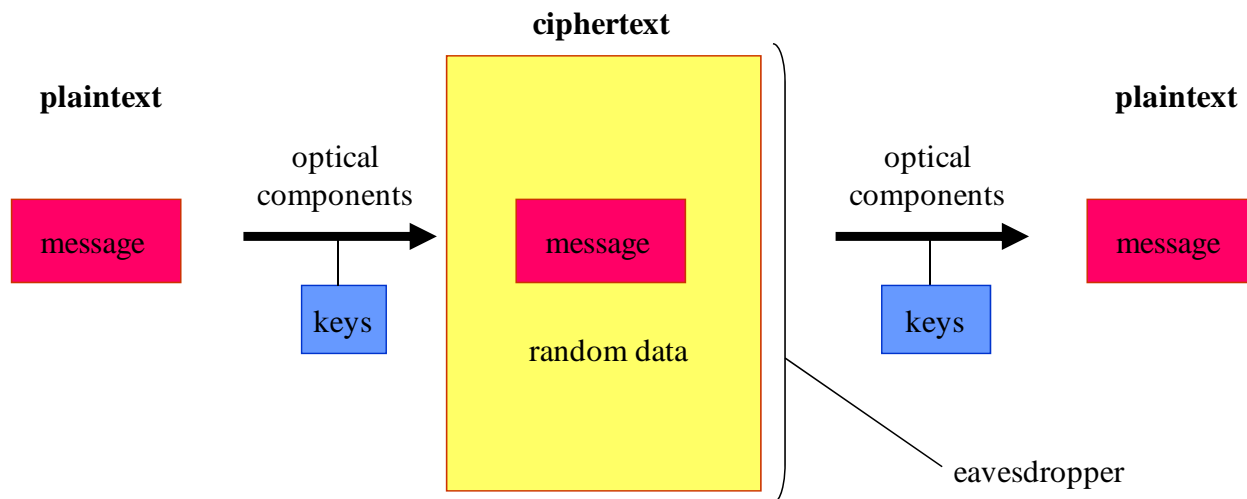


Figure 2 Schematic representation of NIH secure communication

In this project, we investigated what we call the needle-in-the-haystack (NIH) secure communication. It builds upon digital encryption. Messages (Digital ciphertexts) are hidden within randomly generated data. Encryption is performed in physical layer by optical components and so is decryption. The schematic is shown in Figure 2. With correct keys, content other than true message is effectively ignored by intended recipient. But an eavesdropper must record the entirety of message + dummy content, and apply extensive processing to crack

the code. We will show that the attack has to be brute force in nature and requires almost infinite storage capacity and computation time.

2. Optical Implementation of NIH

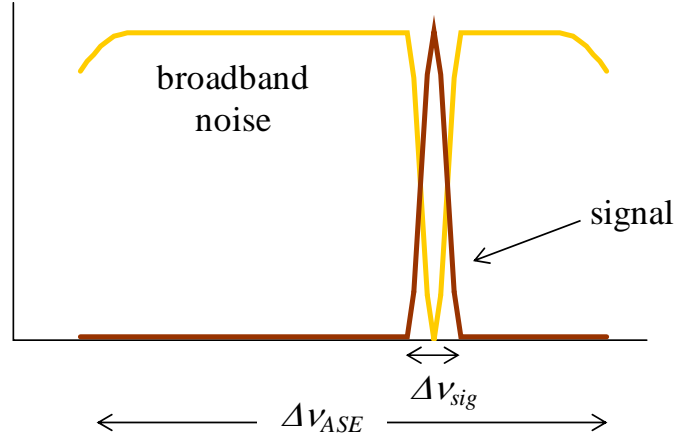


Figure 3 Optical Implementation of NIH

Optical technology is ideally suited for implementation of NIH. Random data (**Haystack**) can be generated using broadband amplified spontaneous emission (ASE) from Erbium-doped fiber amplifiers (EDFA). A notch can be carved in the ASE spectrum using an optical filter. One can then insert the true signal (**Needle**) in its place¹. The Size of haystack is the bandwidth of ASE $\Delta\nu_{ASE}$ and the size of needle is the bandwidth of the signal $\Delta\nu_{sig}$ as shown in Fig. 3.

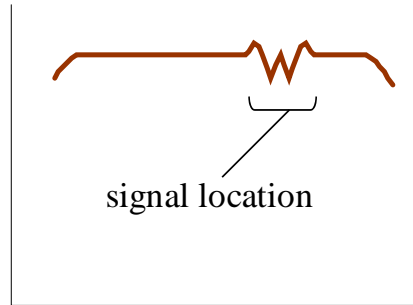


Figure 4 Imperfect stitching betrays signal location in NIH.

This straightforward NIH implementation has an obvious vulnerability: the location of the signal can be determined by observation of the signal in the spectral domain. An eavesdropper can identify signal location by observing frequency channel that deviates from ASE, which is a Gaussian white noise. Any inaccuracy in stitching of signal spectrum into the notch betrays location of true signal channel. Figure 4 shows an example in which the

¹ Pieper et al. (Proc. 29th Southeastern Symp. on System Theory, p. 261-265, 1997) embedded a modulated signal in broadband noise.

notch filter does not fit signal spectrum shape. This method thus requires perfect stitching and leads to impractically tight tolerances in component specifications.

The solution to this vulnerability turns out to be straightforward. Frequency hopping of the notch filter and signal laser in tandem forces eavesdropper to search for signal. The stitching inaccuracies will not make signal visible provided frequency hop interval is short. Intended recipient follows same frequency hop sequence and thus can ignore all random data content.

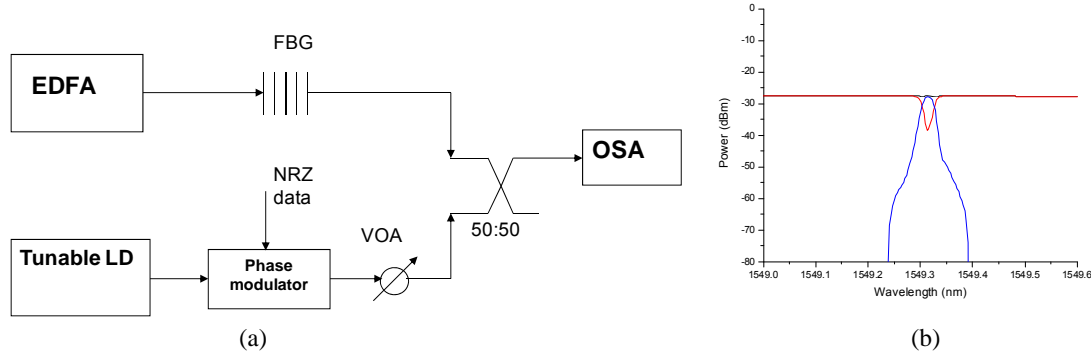


Figure 5 Experimental setup (a) and result (b) for stitching in optical domain.

We have demonstrated that stitching error can indeed be made small so that it is not observable in a certain period of time. In Fig. 5, the observation time is 20 seconds and the stitching is perfect when observed within this time window. As long as frequency hopping rate is greater than once every 20 second, then there is no stitching error because it is below the noise level of the ASE.

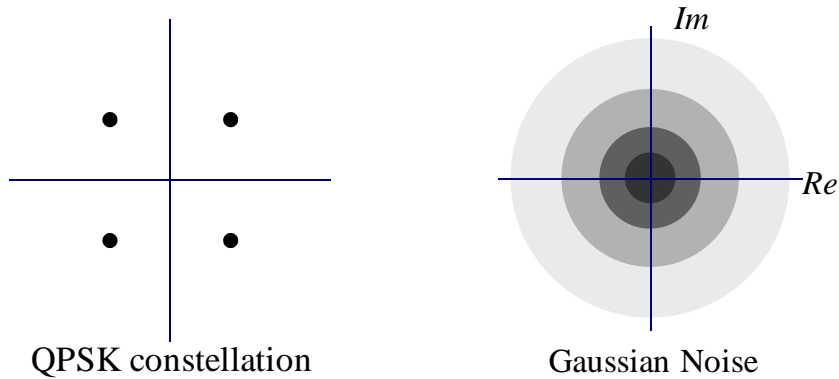


Figure 6 Constellation diagrams of a deterministic signal and ASE

However, there is another vulnerability of the frequency-hopping NIH method. This is because a slice of ASE has Gaussian distribution for both the in-phase & quadrature components. On the other hand, a digital signal inherently takes on a set of discrete values. The constellation diagrams of a deterministic signal and ASE are shown in Fig. 6. An eavesdropper can exploit this difference to identify the true information-bearing signal from the noise.

The information-bearing signal must be disguised to appear like optical noise. The inphase and quadrature parts of one SOP of an optical noise field are each Gaussian noises. Optical noise contains power in both polarization states, although the optical noise could be purposely polarized for this application. The distribution of the inphase part of the information-bearing signal depends on what modulation format is used, but it is typically not Gaussian. For example, if polarization multiplexed QAM is used with M_s bits/symbol, the x -polarization of the QAM signal takes on complex values u_n at the symbol centers ($n = 0,1,2,\dots$), defined by the information to be transmitted. The distribution $f(\text{Re}[u])$ of the real part of u_n comprises $2^{M_s/2}$ delta spikes separated by d . For the example of 16 level QAM, $f(\text{Re}[u])$ comprises four delta spikes.

The first noise-rendering method begins by adding to u_n a pseudorandom complex variable w_n , whose real and imaginary parts each follow a uniform distribution

$$\begin{aligned} f(\text{Re}[w]) &= \frac{1}{d} & -\frac{d}{2} \leq \text{Re}[w] \leq \frac{d}{2} \\ f(\text{Re}[w]) &= 0 & |\text{Re}[w]| > \frac{d}{2} \\ f(\text{Im}[w]) &= \frac{1}{d} & -\frac{d}{2} \leq \text{Im}[w] \leq \frac{d}{2} \\ f(\text{Im}[w]) &= 0 & |\text{Im}[w]| > \frac{d}{2} \end{aligned}$$

The distribution of w_n appears as a square on the complex plane. w_n can be obtained from a pseudorandom data sequence which is generated from the key. The real and imaginary parts of $u_n + w_n$ each has a uniform distribution from $-2^{M_s/2-1}d$ to $2^{M_s/2-1}d$. Then the sum is transformed into a quantity having a Gaussian distribution by taking the inverse error function of each part, and the result can be used for that inphase component of the electric field envelope

$$E_{outx}(n\tau_s) = \sqrt{2} \sigma \text{inverf}\left(\frac{\text{Re}[u_n + w_n]}{2^{M_s/2-1}d}\right) + i \sqrt{2} \sigma \text{inverf}\left(\frac{\text{Im}[u_n + w_n]}{2^{M_s/2-1}d}\right)$$

σ is the standard deviation of a component of the optical noise, to which the signal is to be matched. We have built a simulation tool, schematically shown in Fig. 7, to verify the effective of this noise rendering process. The simulation was carried out using VPI (for PRBS generation and QAM mapping, and coherent optical modulation) and Matlab.

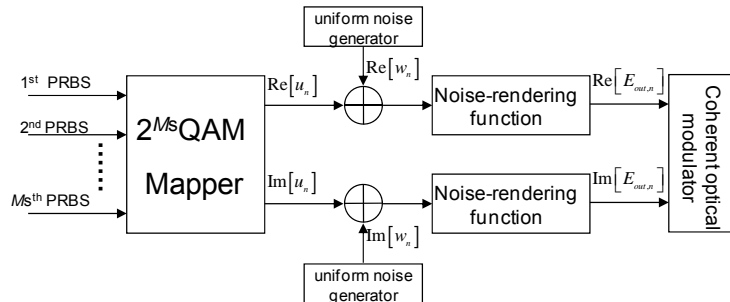


Fig. 7 Block diagram for noise rendering.

Figure 8 presents the constellation diagram of the QPSK signal, the complex noise source, the sum of the above, and the Gaussian noise-like electrical field of the light after noise rendering. The distribution of the real and imaginary part of optical field is clearly Gaussian as shown in Fig. 9.

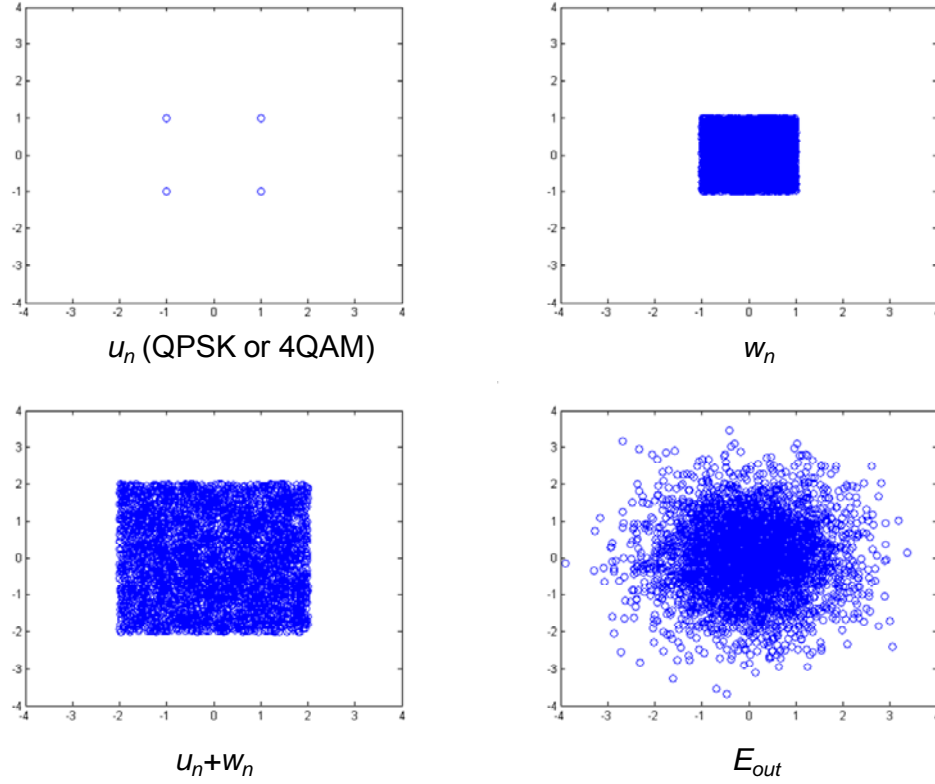


Fig. 8 Constellation diagram of the QPSK signal, the complex noise source, the sum of the above, and the Gaussian noise-like electrical field of the light after noise rendering.

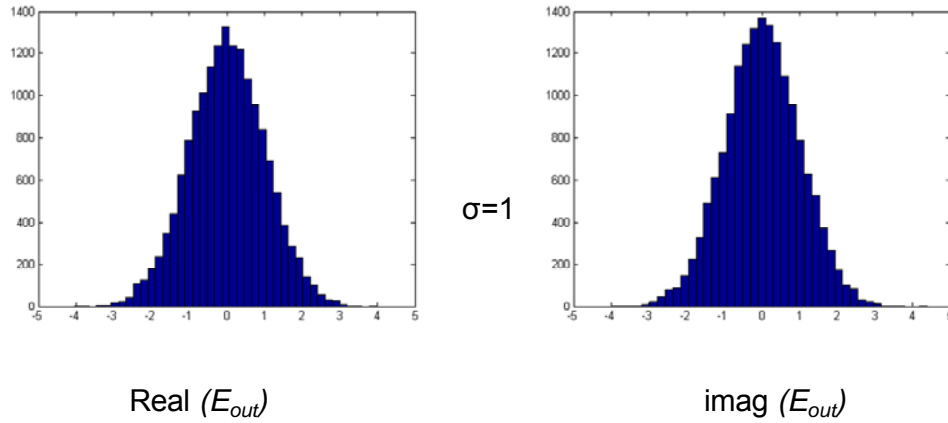


Fig. 9 Distribution of the real and imaginary part of optical field after noise rendering.

There is yet another potential vulnerability associated with frequency-hopping NIH. This is because a slice of ASE is un-correlated, i.e., autocorrelation function is a delta function while

a digital signal inherently has correlation due to limited bandwidth of electrical and optical components used to generated noise-rendered signal as shown in Fig. 10. Potentially, an eavesdropper can exploit this difference to identify the true information-bearing signal from the noise.

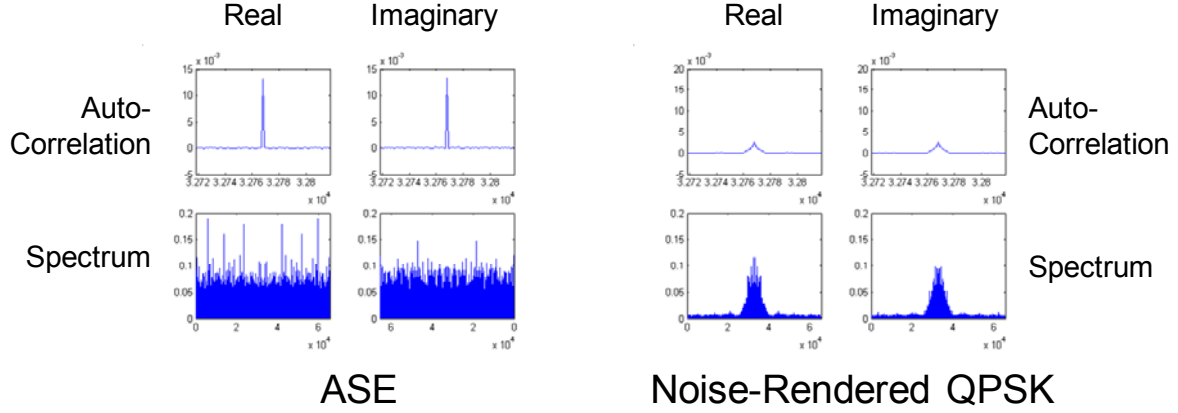


Figure 10 Autocorrelation and spectrum for ASE and noise-rendered QPSK.

Fortunately, this is only a false alarm. Figure 11 shows the autocorrelation function and spectrum for the relevant signals. ASE with notch has anti-correlation in the vicinity of the correlation peak at the center. The autocorrelation of the noise-rendered QPSK signal complements the anti-correlation. As a result, the transmitted signal has the same autocorrelation as the ASE noise.

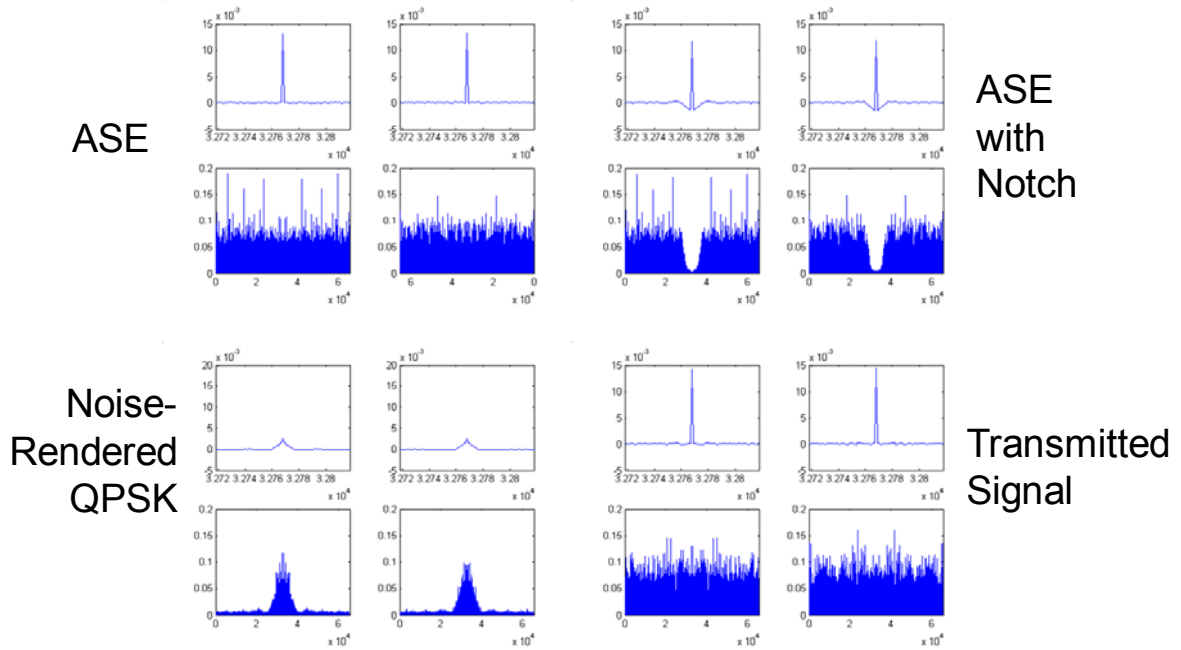


Figure 11 Autocorrelation and spectrum for ASE, ASE with notch, noise-rendered QPSK and transmitted signal.

3. NIH Security

There are two levels of security that the NIH secure communication method affords. The first level is offered in the bounded storage model (BSM). The other level is offered by the computational complexity.

3.1 NIH Security in the BSM

Since the eavesdropper must inspect enormous decrypted ciphertext possibilities to find message, due to frequency hopping and long key in digital encryption, the amount of data to be stored exceeds any reasonable storage capacity available to an eavesdropper. In other words, the eavesdropper cannot afford to even record ciphertext to decrypt later. The concept of presenting eavesdropper with too much data to store is not new², and is called the bounded storage model. Aumann³ proposed publicly sending random number sequence too large to store, then using portions of sequence for encrypted communications such as in satellite broadcast of 200Gb/s random data used by many encrypted communication links.

For the NIH secure communication scheme described in Figure 3, the amount of information to be stored by an eavesdropper can be obtained as follows. Assume $\Delta\nu_{ASE} = 32\text{nm} = 4000\text{GHz}$. The sample rate has to be greater than or equal to $4000\text{GSa/s} \times 2 \text{ quadratures} \times 2 \text{ SOPs} \times 5 \text{ bits resolution}$. This translates to a sample rate of 80Tb/s . For comparison, rate information is stored in the world on all media = 1.3Tb/s , according to a U.C. Berkeley study in 2003.

Therefore it is safe to conclude that the NIH secure communication approach is extremely robust in the bounded storage model.

3.2 Strength of NIH to Brute Force Attack

The NIH optical encryption method ensures that the eavesdropper must use brute force attack to decode. We use a weak digital encryption method namely stream cipher using XOR with a pseudorandom bit sequence as example to find the complexity required for a brute force attack. We assume that the PRBS is generated from a long key, L_{stream} bits. The reason for using stream cipher of a long key length L_{stream} bits is that eavesdropper must have L_{stream} bits correctly received in order to attempt to derive the key. Further, we choose the number of bits in each frequency hop L_{hop} to be much shorter than L_{stream} so that $L_{stream} = M L_{hop}$. The eavesdropper must inspect very many decrypted ciphertext possibilities to find message. The number of possibilities is set by the ratio of random data content to message size. So the number of possibility is determined by optical component parameters and not by amount of digital processing.

² Maurer, J. Cryptol., vol. 5, p. 53-66, 1992

³ IEEE Trans. on Inf. Theory, vol. 48, no. 6, p. 1668-1680, 2002

We will determine the complexity of brute force attack step by step. First consider the case in which the signal is in one frequency channel, encrypted by stream cipher shown in Fig. 12 (a). Assume a linear feedback shift register, length L_{stream} bits, is used to generate the PRBS and the key is the seed of shift register. This digital encryption uses few computations even though key is long. By itself, it is a weak code that is easily broken in time T_{crack} .

Next consider the case in which the signal may appear in any one of $N = \Delta\nu_{ASE} / \Delta\nu_{sig}$ channels, but no frequency hopping shown in Fig. 12 (b). The eavesdropper must inspect $N = \Delta\nu_{ASE} / \Delta\nu_{sig}$ possibilities and the time it takes to break the code would be NT_{crack} .

Finally consider M hops between $N = \Delta\nu_{ASE} / \Delta\nu_{sig}$ channels shown in Fig. 12 (c). To assemble true data content, the eavesdropper must inspect N^M permutations and the time it takes to break the code would be $T_{crack} N^M$. So the NIH encryption scheme effectively amplifies time taken to crack stream cipher N^M .

Now let's quantify the strength of NIH in terms of computation. Take a reasonable values for 10Gb/s signal embedded in EDFA C-band ASE, $\Delta\nu_{ASE} = 32\text{nm} = 4000\text{GHz}$, $\Delta\nu_{sig} = 10\text{GHz}$, $N=400$, frequency hop interval = 100ns, $L_{hop} = 1000$, $L_{stream} = 10000$. This lead to $M=10$ and the number of permutations = $400^{10} = 10^{26}$. Assuming $T_{crack} = 1\mu\text{s}$ (time for simply storing data), the time it take to crack code is 3000 billion years, which is 700 times the age of earth.

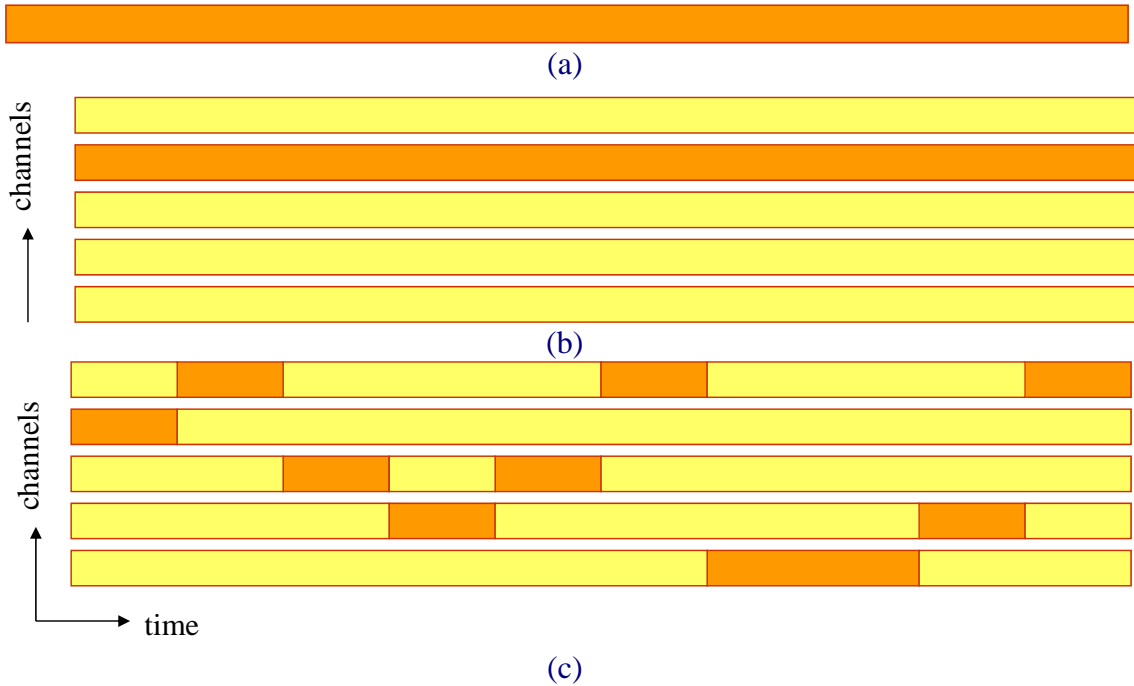


Figure 12 Schematic illustration of permutations in NIH

Therefore, the NIH encryption method resolves the vulnerability of digital encryption methods, that a moderate strength encryption algorithm today is breakable in the future.

4. Conclusion

In this project, we investigated a novel physical layer secure optical communication scheme. The Needle-in-the-Haystack encryption is based on hiding information in random noise. The signal is further rendered so that it is indistinguishable from noise unless a correct key is available. The NIH encryption method offers two levels of security. It is found that, with typical parameters of COTS optical devices, the eavesdropper must have a storage capacity of at least 80 Tb/s and it will take 3000 billion years to crack the code.